EFFICIENCY OF THE MATCHING PROCESS ON THE CZECH REGIONAL LABOUR MARKETS

Daniel Němec, Masaryk University, Faculty of Economics and Administration, Department of Economics

Abstract

This contribution aims to quantify the performance of the Czech regional labour markets and to reveal the most influencing economic factors standing behind. Investigated labour markets are described by the corresponding matching functions. From this point of view the successful matches are treated as an output of production process where unemployed are paired with vacancies. Resulting unemployment outflows are determined by the efficiency of this matching process. Using stochastic frontier model approach, we estimate the efficiency of regional matching functions, evaluate the differences among the regions and reveal the key determinants of this kind of effectiveness. The stochastic frontier is estimated using regional panel data for the period 1997-2013.

Keywords: *matching efficiency, matching function, regional labour markets, stochastic frontier model, panel data, Czech Republic*

JEL Classification: R23, J41, C23, E24
AMS Classification: 62P20

1 INTRODUCTION

Labour market efficiency belongs to the most important factors influencing labour market dynamics and its performance. There are many approaches how to deal with the "efficiency" concept. Most of them are based on the matching function framework which expresses the connection of successful labour market matches as an outcome of interactions between unemployed job seeker and vacancies. This contribution aims to quantify the effectiveness of the Czech labour market from the view of regional labour markets using the stochastic frontier panel data model approach with monthly regional data and explicitly treated fixed effects term in the matching function model equation. On the one hand, this approach extends the previous investigations of the efficiency of the Czech labour market carried out by Němec [5], [6] or by Tvrdoň and Verner [7]. Their results have been based on the aggregate labour market statistics. On the other hand, using the data from monthly regional labour market statistics and stochastic frontier panel data model methodology, it offers a new insight into the outcomes of the Czech labour market in the last 15 years and extends the detailed analysis of Galuščák and Münich [2] in a specific way, dealing with efficiency issues.

Stochastic frontier model approach has been used by Ilmakunnas and Pesola [4] in their study of regional labour markets in Finland. They used annual data and did not take into account explicitly possible individual fixed effects of examined regions. Gorter et al. [3] investigated the efficiency in the Dutch labour market in Netherland along the same lines. They have observed that the estimated labour market efficiency increases during the recession and recovery period while it decreases during the economic booms. This interesting feature is considered in this contribution as well.

2 STOCHASTIC FRONTIER MODEL WITH PANEL DATA

Stochastic frontier model approach allows us to measure the performance of production units which use inputs to produce outputs of any kind. Production technology is described by the production function. This parametric approach to measure technical inefficiency may be used in many applications. As for the labour market framework, the production technology of a labour market is usually described by the matching function.

2.1 Matching function and matching efficiency

The matching function expresses the interaction mechanism between the unemployed and vacancies. This concept is based on the fact that both the flows of unemployed and the flows of unfilled job vacancies are able to meet each other. This dynamic relationship could be described simply by a standard production function with two inputs: the unemployed and the vacancies. New matches are thus an outcome of this matching process. In my contribution, the regional labour markets are represented by a standard Cobb-Douglas matching function in log-linear form:

$$\ln M_{it} = \alpha_i + \beta_u \ln U_{it} + \beta_v \ln V_{it} + \varepsilon_{it} , \qquad (1)$$

where $i = 1,\ldots,N$ denotes the regions and $t = 1,\ldots,T$ the time periods. The α_i terms are fixed regional effects and ε_{it} represents stochastic factors. This basic form of matching function may be extended and modified in many ways. Ilmakunnas and Pesola [4] implemented regional and labour force characteristics directly into the matching function by means of other explanatory variables. Resulting efficiency was thus a linear function of regional fixed effects and various regional characteristics. In their view, the term ε_{it} was treated purely as white noise process. Similar approach may be found in the work of Gorter et al. [3]. Galuščák and Münich [2] enhanced the basic matching function form by the flow factors (i.e. unemployment and vacancy inflows realized during the time period). Stochastic frontier model approach tries to model the stochastic term ε_{it} as consisting of combination of random variations in the matching process and the region specific inefficiency term. Regional and labour force characteristics are then implemented directly into this inefficiency term. This approach was used by Ilmakunnas and Pesola [4]. But they did not include the fixed (or random) region effects. In my contribution, I try to estimate the inefficiency of the Czech regional labour market using fixed effect panel stochastic model. This model approach is thus able to capture region specific individual effects, basic matching function characteristics and time-varying regional inefficiency terms at once.

2.2 Fixed effect panel stochastic model

To estimate matching function parameters and the inefficiency of the matching process we use the approach proposed by Wang and Ho [8]. Their specification of a stochastic frontier model is as follows:

$$y_{it} = \alpha_i + \mathbf{x_{it}\beta} + \varepsilon_{it}, \qquad (2)$$

$$\varepsilon_{it} = v_{it} - u_{it}, \qquad (3)$$

$$v_{it} \sim N\left(0, \sigma_v^2\right), \qquad (4)$$

$$u_{it} = h_{it} \cdot u_i^*, \qquad (5)$$

$$h_{it} = f\left(\mathbf{z_{it}\delta}\right), \qquad (6)$$

$$u_i^* \sim N^+\left(\mu, \sigma_u^2\right), \qquad i = 1,\ldots,N \qquad t = 1,\ldots,T . \qquad (7)$$

In this model framework, α_i is individual fixed effect for the unit i, $\mathbf{x_{it}}$ is a $1 \times K$ vector of explanatory variables, v_{it} is a random error with zero mean, u_{it} is a stochastic variable measuring inefficiency, and h_{it} is a positive function of a $1 \times L$ vector of non-stochastic determinants of inefficiency ($\mathbf{z_{it}}$). Constant term is excluded from explanatory variables and inefficiency determinants. It should be clear that the notation N^+ means that the realized values of the variable u_i^* are positive. In case of $\mu = 0$ the variable u_i^* follows a half-normal distribution.

Wang and Ho [8] showed how to remove the fixed individual effect from the model. This procedure allows us to estimate all the model parameters. Of course, the individual effects may be recovered from the final parameter estimates. There are two possible approaches to model transformation: first-differencing and within-transformation. Both methods are equivalent (see Wang and Ho [8]). Stochastic frontier model of the Czech regional labour markets has been identified using the first-difference transformation. The main points of this methodology may be described as follows (for detailed discussion see Wang and Ho [8]).

It is necessary to define first difference of corresponding variables as $\Delta w_{it} = w_{it} - w_{it-1}$ and the stacked vector of Δw_{it} for a given i and $t = 2,\ldots,T$ is denoted as $\Delta \tilde{w}_i = (\Delta w_{i2}, \Delta w_{i3}, \ldots, \Delta w_{iT})'$. Assuming that the function h_{it} is not constant, i.e. the vector z_{it} contains at least one time-varying variable, the model in its first-difference form may be expressed as:

$$\Delta \tilde{y}_{it} = \Delta \tilde{\mathbf{x}}_{\mathbf{it}} \boldsymbol{\beta} + \Delta \tilde{\varepsilon}_i, \tag{8}$$

$$\Delta \tilde{\varepsilon}_i = \Delta \tilde{v}_i - \Delta \tilde{u}_i, \tag{9}$$

$$\Delta \tilde{v}_i \sim MN(0, \Sigma), \tag{10}$$

$$\Delta \tilde{u}_i = \Delta \tilde{h}_i u_i^*, \tag{11}$$

$$u_i^* \sim N^+(\mu, \sigma_u^2), \qquad i = 1,\ldots,N \tag{12}$$

It is obvious from panel data models that first-difference introduces correlations of Δv_{it} within the ith panel. The covariance matrix of the multivariate distribution of $\Delta \tilde{v}_i$ is

$$\Sigma = \begin{bmatrix} 2\sigma_v^2 & -\sigma_v^2 & 0 & \cdots & 0 \\ -\sigma_v^2 & 2\sigma_v^2 & -\sigma_v^2 & \cdots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\sigma_v^2 \\ 0 & 0 & \cdots & -\sigma_v^2 & 2\sigma_v^2 \end{bmatrix}. \tag{13}$$

The covariance matrix Σ has elements $2\sigma_v^2$ on the diagonal and $-\sigma_v^2$ on the off-diagonal. The key point revealed by Wang and Ho [8] is that the distribution of the term u_i^* is unaffected by the transformation. This fact helps to derive marginal log-likelihood function for each panel unit:

$$\ln L_i = -\frac{1}{2}(T-1)\ln(2\pi) - \frac{1}{2}(T-1)\ln(\sigma_v^2) - \frac{1}{2}\Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{\varepsilon}_i + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right) \tag{14}$$

where

$$\mu_* = \frac{\mu/\sigma_u^2 - \Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{h}_i}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2} \qquad \sigma_*^2 = \frac{1}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2} \qquad \Delta\tilde{\varepsilon}_i = \Delta\tilde{y}_i - \Delta\tilde{\mathbf{x}}_\mathbf{i}\boldsymbol{\beta}. \tag{15}$$

In this expression, Φ is the cumulative density function of a standard normal distribution. Log-likelihood function of the model is obtained by summing the above function over al panel units.

Wang and Ho [8] approximated the observation specific technical inefficiency as conditional expectation

$$E(u_{it}|\Delta\tilde{\varepsilon}_i) = h_{it}\left[\mu_* + \frac{\phi(\mu_*/\sigma_*)\sigma_*}{\Phi(\mu_*/\sigma_*)}\right] \tag{15}$$

evaluated at estimated values of term $\Delta\tilde{\varepsilon}_i$. This is a modified estimator of inefficiency terms which uses $\Delta\tilde{\varepsilon}_i$ instead of $\tilde{\varepsilon}_i$ as the conditional term. The main advantage is that the vector $\Delta\tilde{\varepsilon}_i$ contains all the information of individual unit in the sample and does not depend on individual effect term α_i that has the variance of higher order in case of small time dimension of the sample (variance of order $1/T$ in comparison to the variance of $1/((N-1)T)$ for the estimator $\hat{\beta}$). Technical efficiency may be obtained in accordance with other studies (see

Battese and Coelli [1]) as $\exp(-u_{it})$. For derivation of individual fixed effects terms see Wang and Ho [8].

2.3 Data and model specification

The model for the Czech regional labour markets is estimated using the monthly data set covering a sample of 77 districts from the January 1997 to the June 2013. In comparison with the other authors I try to use this "high" frequency data set due to fact that the aggregation may lead to some losses of information. Galuščák and Münich [2] worked with quarterly data, Ilmakunnas and Pesola [4] and Gorter et al. [3] focused on annual data of regions in Finland and Netherland respectively.

The original data come from database of the Ministry of Labour and Social Affairs which cover the data from regional Employment offices. I used the following variables: the number of registered successful matches, M_{it}, in the corresponding month, the number of unemployed at the start of the month, U_{it}, and the number of vacancies at the start of the month, V_{it}. All the data are seasonally unadjusted because we treat the seasonal patterns of the labour market characteristics within the factors influencing the inefficiency term.

Panel data set consists of 77 districts and 198 monthly periods. In accordance with Galuščák and Münich [2], three districts were omitted: Praha, Praha-Východ and Praha-Západ. These labour markets are too specific in their labour market dynamics. The estimated model has the form defined by the equations (2)-(7), where $y_{it} = \ln M_{it}$, $\mathbf{x_{it}} = (\ln U_{it}, \ln V_{it})$, $\boldsymbol{\beta} = (\beta_u, \beta_v)'$, $h_{it} = \left|\delta_{time}t + \delta_{time^2}t^2 + \delta_{Q2}Q2 + \delta_{Q3}Q3 + \delta_{Q4}Q4\right|$, where t represents the time trend and $Q2$, $Q3$ and $Q4$ are seasonal quarterly dummies for the 2nd, 3rd and 4th quarters respectively. I have defined $\mu = 0$ and we have thus a half-normal representation of the model. There is no possibility to obtain district specific labour market characteristic due to monthly data frequency. Inefficiency terms capture time trend (which is usual in many applications, see e.g. Battese and Coelli [1]) and seasonal factors which may be thus connected directly with the efficiency of the labour markets. For computational purposes, the variance parameters were parameterised as $c_v = \ln\sigma_v^2$ and $c_u = \ln\sigma_u^2$ respectively.

3 EFFICIENCY ESTIMATES

The model parameters were estimated by numerically maximizing the sum of marginal log-likelihood functions (13). All the estimation procedures were performed using Matlab version 2013b and its implemented function for unconstrained maximization. I have estimated two model specifications. The first model was estimated using the full sample of the period 1998 to 2013. To reveal the possible changes in the model parameters and to capture the time variations of inefficiency in more detail, the yearly rolling estimates were carried out. Separate model were thus estimated for the years 1998-2012. In this case, the quadratic term in the inefficiency scaling factor h_{it} was omitted.

Table 1: Parameter estimates (full sample 1997-2013)

$\beta_{\log(u)}$	$\beta_{\log(v)}$	δ_{time}	δ_{time^2}	δ_{Q2}	δ_{Q3}	δ_{Q4}	$\log\sigma_v^2$	$\log\sigma_u^2$
0.7131	0.0896	0.2843	0.0652	-0.2708	0.1102	0.3794	-2.6163	-1.1772

Table 1 presents the estimated parameters for the model covering the whole sample period. Estimated coefficients $\beta_{\log(u)}$ and $\beta_{\log(v)}$ does not confirm the empirical findings that with regional data it may be more likely to find increasing returns in matching (see Ilmakunnas and Pesola [4]). But, as it will be seen later, this conclusion may be an outcome of strong assumption that these parameters remain stable through the whole period. On the other hand, the tendencies provided by the estimates of parameters in the efficiency scale factor h_{it} are evident. Regarding the parameters on the time trend variables we can see the rising inefficiency patterns in the regional matching processes. This negative development is typically reverted in the second quarter of each year. Higher variability, σ_u^2, of the inefficiency term in comparison to the white noise process variability, σ_u^2, contributes to the satisfying identification of the stochastic frontier model (as stated by Wang and Ho [8]).

Figure 1 shows the interquartile range of inefficiency terms distributions for all 74 districts using the estimates on full sample period. The number corresponds with the sorting order of districts in the source data files provided by Ministry of Labour and Social Affairs (after excluding the Prague regions).

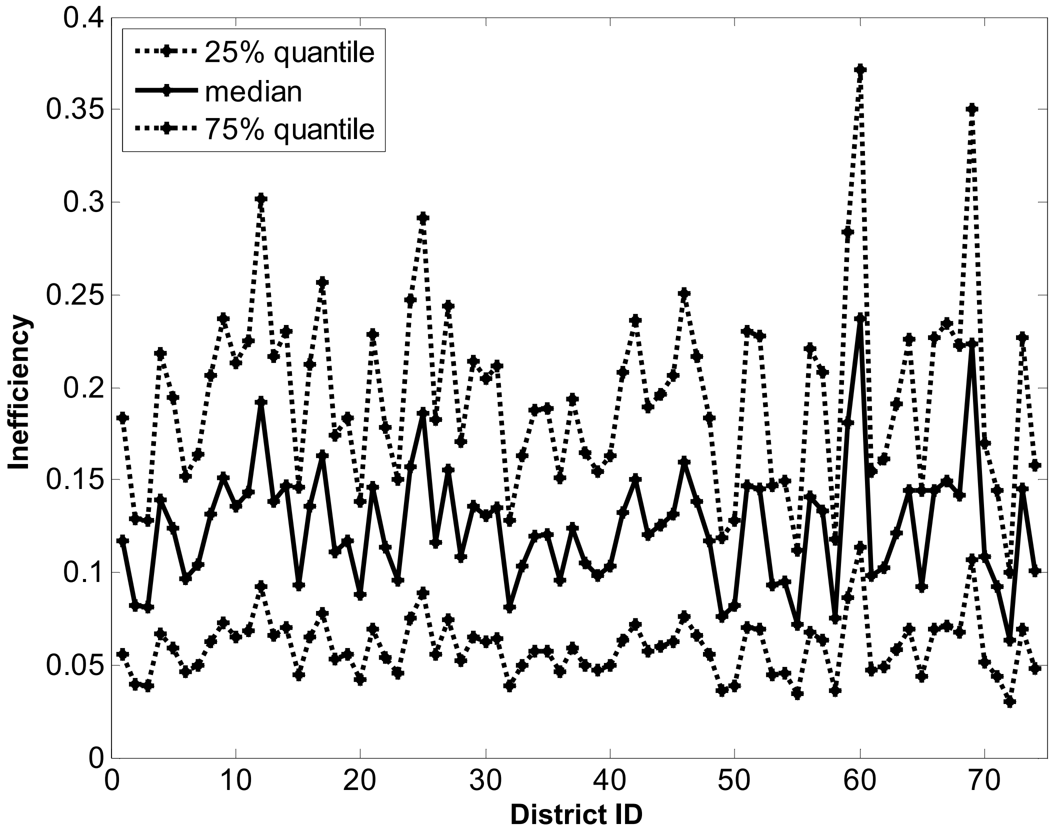

Figure 1: Inefficiency range (full sample 1997-2013)

Minimum inefficiency values for each district (which are not presented here) are almost zero for all investigated labour markets. It is this clear that all the regions are able to match the unemployed with the vacancies at the full rate. It is caused mostly by the seasonal factors in the second quarter of the year. Figure 1 suggests that there are some regions with exceptionally good or bad efficiency performance.

A detailed view on the inefficiency patterns in selected districts may be found in the Table 2. It may be surprising that one of the best performing districts is the district Most or Nový Jičín. These districts do not belong to the regions with the low unemployment properties. But, it should be stressed that low inefficiency does not automatically mean low unemployment. It expresses the potential for new created matches which can be constituted by the interaction between unemployed and available vacancies.

Table 2: Selected district inefficiency patterns (full sample 1997-2013)

ID	District	Minimum	25% quantile	50% quantile	75% quantile	Maximum
2	Beroun	0.0006	0.0393	0.0821	0.1288	0.2274
32	Most	0.0006	0.0390	0.0816	0.1280	0.2260
72	Nový Jičín	0.0004	0.0304	0.0636	0.0997	0.1761
25	Cheb	0.0013	0.0889	0.1858	0.2915	0.5146
60	Jeseník	0.0016	0.1132	0.2367	0.3713	0.6556
69	Bruntál	0.0015	0.1068	0.2231	0.3501	0.6181

From this point of view, these results imply that the potential of labour market is utilized quite well. There may be an appropriate structure of unemployed and vacancies, unobserved characteristics of the unemployed support their willingness to active job search and finally, the surrounding regions may offer other possibilities for employing unemployed job applicants (this spatial dependency is not implemented in estimated models so far). The unfavourable efficiency outcomes of the districts Jeseník or Bruntál may be thus explained in a similar way.

Table 3: Parameter estimates (rolling windows 1998-2012)

	$\beta_{\log(u)}$	$\beta_{\log(v)}$	δ_{time}	δ_{Q2}	δ_{Q3}	δ_{Q4}	$\log\sigma_v^2$	$\log\sigma_u^2$
1998	1.3574	0.2712	0.0960	-0.3840	-0.3393	-0.2636	-2.8048	-1.1863
1999	1.7943	0.0542	-0.1082	0.4329	0.5805	0.6225	-3.0978	-1.2216
2000	1.5210	0.1923	-0.0584	0.2336	0.3001	0.3031	-3.0569	-1.2847
2001	1.8825	0.4025	0.1182	-0.4728	-0.5923	-0.6603	-3.0326	-1.1917
2002	0.5865	0.2363	0.0489	-0.1958	0.0204	0.0687	-2.8258	-1.2155
2003	0.6480	0.1730	0.0419	-0.1678	-0.0533	0.0364	-3.0108	-1.2502
2004	2.4917	0.1851	0.1133	-0.4723	-0.7019	-0.7737	-3.0411	-1.1916
2005	0.6432	0.1400	0.0540	-0.2161	0.0076	0.0256	-2.7998	-1.2485
2006	1.8826	0.3914	0.1608	-0.6820	-0.8764	-1.2779	-2.8681	-1.0939
2007	0.6849	0.3607	0.0566	-0.2263	0.1060	0.0012	-2.6297	-1.2139
2008	0.6581	0.5464	0.0481	-0.1923	0.1161	-0.0104	-2.7926	-1.2154
2009	2.1295	0.1837	0.1054	-0.4215	-0.4260	-0.3535	-2.9602	-1.3281
2010	0.5669	0.4198	0.0700	-0.2799	0.0156	0.1354	-2.4649	-1.2040
2011	1.1749	0.3948	0.1437	-0.5749	-0.6936	-0.7646	-2.7806	-1.1537
2012	0.8342	-0.0053	0.1811	-0.7245	-0.6525	-0.4728	-2.6561	-0.9824

Table 3 shows the changes in point estimates of model parameters based on the estimates using the yearly rolling window. In this case, there are years with high match elasticity to unemployed. This feature leads naturally to the increasing returns to scale which is in accordance with prevailing literature dealing with regional labour market data. The negative elasticity of matching outcomes to the vacancies should be treated as zero. This is a sign of worsening labour market conditions in the Czech economy. Due to shorter time span, only the linear trend variable has been used in the inefficiency scaling parameter h_{it}.

The inefficiency within the year tends to be rising and accompanied by important seasonal patterns, especially by positive effect on the matching function outcomes in the second quarters. Results from the Table 3 highlight the needs to incorporate possible parameter changes into the modelling procedures. Another explanation of the parameter instability may be the lack of regional specific inefficiency variables varying across the time and cross-sections. That remains as an important task for further model enhancements.

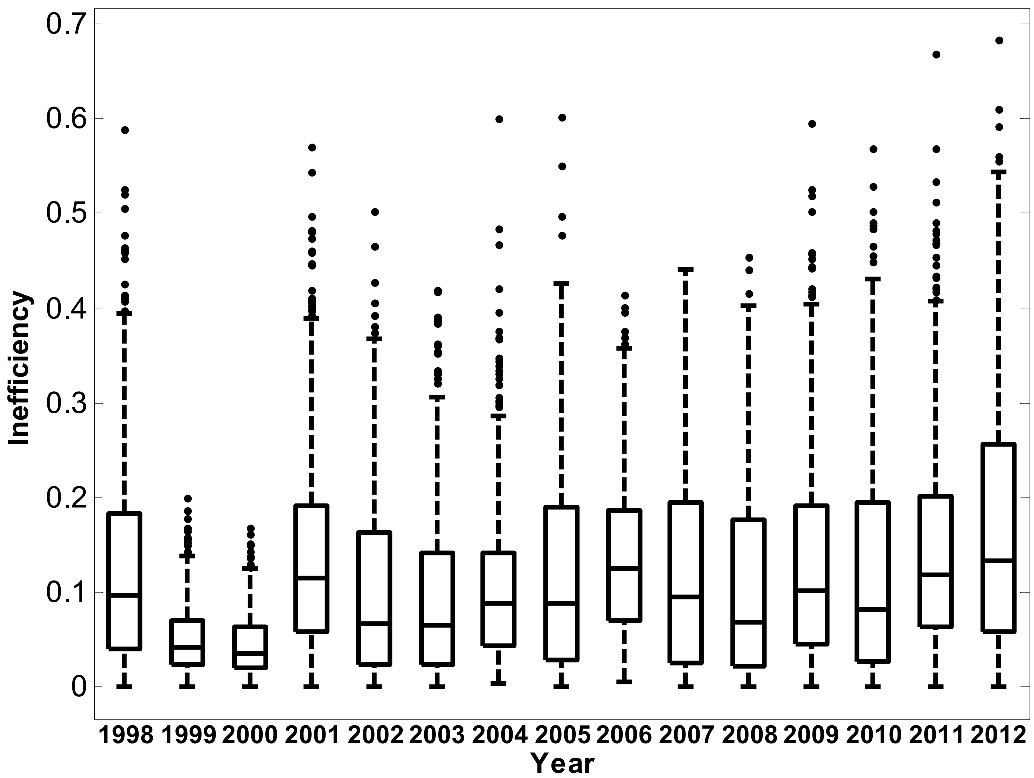

Figure 2: Inefficiency distributions (rolling windows estimates 1998-2012)

Figure 2 depicts the distribution of inefficiency terms across the Czech districts during the period from 1998 to 2012. This figure summarise the aggregate regional inefficiency changes in a straightforward way. We can observe the periods of 1999 and 2000 performing low differences in the efficiency of the regional labour markets. It is the period after the economic crisis of 1997. These years may be described by rising unemployment rates in all regions. But, it seems that the rise of unemployment was accompanied in general by the effective vacancy posting.

The differences across the regions started to be diminishing in the period from 2001 to 2004. After 2005 the variability in inefficiency properties of the regional labour markets tends to be rising again. We may observe the biggest diversity in the 2012. These results do not indicate that the estimated labour market inefficiency may rise during the recession and recovery period while it decreases during the economic booms. Regarding the fact that Gorter et al. [3] used annual data, it should be noted that this contradiction is not conclusive.

4 CONCLUSION

In my contribution, I have presented an alternative approach to measure the efficiency of the matching process on the Czech regional labour markets. Obtained results shows, that the stochastic frontier model approach is able to capture some interesting patterns of these labour markets controlling individual fixed effects of examined districts and possible time-varying changes in the inefficiency terms. The model estimates using the full sample displays rising tendency of matching inefficiency in all districts with strong seasonal patterns. These tendencies are accompanied by rising disparities among the regions although the low inefficiency does not necessary mean the low unemployment in the investigated districts.

It will be of great importance in further research to focus on the model outcomes using the aggregate quarterly and yearly data that allows including region specific variables. Moreover, the spatial properties of the labour markets dynamics should be investigated, i.e. the efficiency terms should incorporate the influence of neighbouring districts.

Acknowledgements: This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0808/2013.

References

[1] Battese, G. E., and Coelli, T. J.: A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* **20** (1995), 325–332.

[2] Galuščák, K., and Münich, D.: Structural and Cyclical Unemployment: What Can Be derived from the Matching Function. *Czech Journal of Economics and Finance* **57** (2007), 102–125.

[3] Gorter, C., Nijkamp, P., and Pels, E.: Vacancy Dynamics and Labor Market Efficiency in the Dutch Labor Market. *Growth and Change* **28** (1997), 173–200.

[4] Ilmakunnas, P., and Pesola, H.: Regional Labour Market Matching Functions and Efficiency Analysis. *Labour* **17** (2003), 413–437.

[5] Němec, D.: Evaluating labour market flexibility in V4 countries. In: *Proceedings of 31st International Conference Mathematical Methods in Economics* (Vojáčková, H., ed.). College of Polytechnics Jihlava, Jihlava, 2013, 661–666.

[6] Němec, D.: Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *The Czech Economic Review* **7** (2013), 21–41.

[7] Tvrdoň, M., and Verner, T.: Regional unemployment disparities and their dynamics: Evidence from the Czech Republic. In: *Proceedings of 30th International Conference Mathematical Methods in Economics* (Ramík, J., and Stavárek, D., eds.). Silesian University in Opava, School of Business Administration in Karviná, 2012, 938–943.

[8] Wang, H., and Ho, C.: Estimating fixed-effect panel stochastic frontier models by transformation. Journal of Econometrics 157 (2010), 286–296.

Author's address

Ing. Daniel Němec, PhD.

Masaryk University, Faculty of Economics and Administration, Department of Economics

Lipová 41a, 602 00 Brno

Czech Republic

email: nemecd@econ.muni.cz